

上海大学出版社

2006年上海大学博士学位论文 3



模糊概念格模型 及其应用研究

- 作者：强 宇
- 专业：控制理论与控制工程
- 导师：刘宗田



上海大学出版社

2006年上海大学博士学位论文 3



模糊概念格模型 及其应用研究

- 作者：强 宇
- 专业：控制理论与控制工程
- 导师：刘宗田



上海大学

本论文经答辩委员会全体委员审查,确认符合上海大学博士学位论文质量要求。

答辩委员会名单:

主任:	胡运发	教授,复旦大学计算机系	200433
委员:	黄国兴	教授,华东师范大学软件学院	200062
	周傲英	教授,复旦大学计算机系	200433
	苗春霖	教授,同济大学计算机学院	200092
	李 青	教授,上海大学计算机学院	200072
导师:	刘宗田	教授,上海大学计算机学院	200072

评阅人名单:

胡运发	教授,复旦大学计算机系	200433
黄国兴	教授,华东师范大学软件学院	200062
周傲英	教授,复旦大学计算机系	200433

评议人名单:

刘晚平	教授,合肥工业大学计算机学院	230009
汤庸	教授,广州中山大学信息科学技术学院	510275
苗夺谦	教授,同济大学计算机学院	200092
李青	教授,上海大学计算机学院	200072
彭智勇	教授,武汉大学软件学院	430072

答辩委员会对论文的评语

模糊概念格模型及其应用研究是前沿性研究课题。该选题具有重要的理论价值和潜在实际意义。

该论文的主要贡献是将模糊理论与形式概念分析相结合,较深入地研究了基于连续数值模糊集的模糊概念格模型及在知识发现中的应用。研究成果在以下几点上有所创新:

(1) 提出了一种模糊概念格模型,定义了两个模糊参数 E 和 δ , 分别是概念的外延中对象的平均隶属度和对象的隶属度值相对平均隶属度的偏离程度,使之具有更广泛应用。

(2) 提出了用渐进方法构造模糊概念格的算法,在该算法中采用了保留中间结果 k_d, h_d 的方法渐进式计算模糊参数 E, δ ,使之不需要回到初始背景中计算。

(3) 提出了基于模糊概念格的模糊关联规则的提取方法与算法,其中利用模糊参数 δ ,可以避免生成非健壮节点对,从而防止生成效果欠佳的规则,实验并分析了算法的有效性和时空复杂度。

(4) 针对模糊概念格构造的巨大的时空复杂度问题,将分布式处理的思想引入模糊概念格,提出了分布式模糊概念格的模型,并且基于模糊概念子格的并运算,提出并实验分析了分布式构造模糊概念格的算法。此算法比未采用分布式处理在时空复杂度上有很大改进。

论文反映出作者已较全面地掌握了与本课题相关的国内外发展动态,且具有扎实的基础理论及系统深入的专业知

识,显示了该生已具有相当强的独立科研能力。论文立论正确,论据充分,条理性好,层次分明,实验数据可靠。在答辩过程中,回答问题思路清晰,能予以正确回答。

答辩委员会表决结果

经答辩委员会表决,全票同意通过强宇同学的博士学位论文答辩,建议授予工学博士学位。

答辩委员会主席: **胡运发**

2006 年 4 月 26 日

摘 要

随着计算机技术的飞速发展,数字信息与日俱增,信息存储和处理任务越来越艰巨和困难,因此,对信息表示和处理的研究具有非常重要的意义。在实际应用中,信息多是模糊的、不确定的,故模糊信息的表示及处理研究具有重要的实用意义。

作为一种优良的数学工具,概念格已经广泛应用于知识表示、数据挖掘、信息检索等许多领域。在多数文献中,概念格的研究是基于标准形式背景的。但在实际中,信息多是模糊的、不确定的。目前关于模糊信息的研究方法多数是基于有限的 L -模糊集合的,即基于离散数值的模糊集合,无法表示连续数值的模糊信息。

本文所做的工作是将模糊理论与形式概念分析结合,研究了基于连续数值模糊集合的模糊概念格模型及在知识发现中的应用。

本文的主要工作和研究成果具体是:

1. 提出了一种模糊概念格模型,该模型基于连续数值模糊集合,具有更广泛的应用。在模糊概念格的节点级上,定义了两个模糊参数 E 和 δ ,分别是概念的外延中对象的平均隶属度和对象的隶属度值相对平均隶属度的偏离程度。

-
- 国家自然科学基金(60275022)
 - 国家自然科学基金(60575035)
 - 国家自然科学基金(69985004)

2. 提出并实现了一种用渐进式方法构造模糊概念格的算法,在该算法中采用了保留中间结果 k_d, h_d 的方法渐进式计算模糊参数 E, δ ,使之不需要回到初始背景中计算。

3. 提出了基于模糊概念格的模糊关联规则的提取方法与算法,其中利用模糊参数 δ ,可以避免生成非健壮节点对,从而防止生成效果欠佳的规则,实验并分析了算法的有效性和时空复杂度。

4. 针对模糊概念格构造的巨大的时空复杂度问题,将分布式处理的思想引入模糊概念格,提出了分布式模糊概念格的模型,并且基于模糊概念子格的并运算,提出并实验分析了分布式构造模糊概念格的算法。此算法比未采用分布处理在时空复杂度上有很大改进。

5. 提出了采用模糊聚类技术和聚类参数从模糊概念格生成模糊聚类、构造模糊概念层次,进而采用映射构造模糊本体的方法,并给出了实例。

关键词 模糊形式概念分析,模糊概念格,模糊关联规则,分布式模糊概念格,模糊聚类,模糊本体。

Abstract *

With the rapid development of computer science, and the rapid increase of digital information, information storage and process task is more and more difficult, so research on information representation and process has very important significance. In application, information is fuzzy and indeterminate, so research on fuzzy information representation and transaction has extremely important significance in theory and application.

As a kind of excellent mathematic tool, Concept Lattice has been widely applied in knowledge representation, datamining and many other fields. In most papers, Concept Lattice is based on standard context, but in practice, information is usually fuzzy and indeterminate.

At present, most of research methods on fuzzy information are based on L -fuzzy set, they can't process the fuzzy information of consecutive value.

In the dissertation, we combine fuzzy information with Concept Lattice, research the fuzzy concept lattice model

* This work is supported by the National Natural Science Foundation of China (60275022), (60575035) and (69985004).

based on consecutive value, and study it's application in knowledge discovery.

The main work and research results in this dissertation are as follows:

1. We present a model of Fuzzy Concept Lattice, this model is based on fuzzy set of consecutive value , so it has more wide application. At the node level of Fuzzy Concept Lattice, we introduce two fuzzy parameters E and δ , which are respectively mean value of membership degree of the objects in extent of the fuzzy concept and deviation degree of membership degree of the objects for the mean value.

2. We present and implement an algorithm for generating Fuzzy Concept Lattice by incremental method. In this algorithm, we employ two intermediate variables to compute fuzzy parameters E and δ incrementally, so that we needn't to return into the initial context to compute them.

3. We present a method and an algorithm for fuzzy association rule extraction based on Fuzzy Concept Lattice. In the method we apply fuzzy parameter δ to avoid creating non-robust node parts, and thus avoid creating impotent rule. An experiment is completed, and the effect and time-spatial complexity of the algorithm is analysed.

4. Considering the large time and spatial complexity of lattice construction, we introduce the thought of distributed process into Fuzzy Concept Lattice, present a model of

distributed Fuzzy Concept Lattice, and the union operation based on Fuzzy Concept Sub-Lattice. An experiment for an algorithm of Distributed generating Fuzzy Concept Lattice is completed and results is analysed. The results show that the algorithm is superior to those algorithms without distributed process on time-spatial complexity.

5. We apply fuzzy cluster techniques and the fuzzy parameters to create fuzzy cluster from Fuzzy Concept Lattice, construct fuzzy concept hierarchy, and apply mapping method to construct fuzzy ontology. An experimental case and analysis are given.

Key words fuzzy concept lattice, fuzzy parameters, fuzzy associate rule, distributed fuzzy concept lattice, fuzzy cluster, fuzzy ontology.

目 录

第一章 绪论	1
1.1 课题研究的目的和意义	1
1.2 国内外研究概况	1
1.3 论文的主要研究内容	2
第二章 概念格和模糊集基本理论	4
2.1 概念格基本理论	4
2.1.1 基本定义	4
2.1.2 概念定标与逻辑定标	5
2.1.3 概念知识处理	5
2.1.4 概念格结构在计算机各领域的应用	6
2.2 模糊集理论	8
2.2.1 模糊集分解理论	9
2.2.2 模糊集扩展定理	9
第三章 模糊概念格模型及其构造	11
3.1 引言	11
3.1.1 Karl Erich Wolff 的模糊概念表示	11
3.1.2 Burusco 计算 L -模糊概念格的方法	15
3.1.3 RALAMBONDRAINY 的模糊信息表示	17
3.1.4 GIRARD 和 Heari RALAMBONDRAINY 的 模糊格法	18
3.2 模糊概念格数学模型	22
3.3 模糊概念格的构造算法	25

3.4	实验及分析	29
3.5	小结	31
第四章	基于模糊概念格的模糊关联规则提取算法研究	33
4.1	引言	33
4.1.1	内涵缩减的定义及性质	34
4.1.2	族集的最小覆盖	36
4.1.3	计算族集最小覆盖集算法	37
4.1.4	计算内涵缩减集	37
4.1.5	基于内涵缩减的规则提取	38
4.2	基于健壮节点对的模糊关联规则生成	38
4.2.1	模糊参数 E, δ	39
4.2.2	模糊关联规则形式定义	39
4.3	模糊关联规则生成算法及实验分析	39
4.3.1	基于健壮节点对的提取算法	39
4.3.2	基于内涵缩减的提取算法	45
4.4	模糊规则提取在 TSP 度量模型中的应用	48
4.4.1	数据处理	49
4.4.2	规则挖掘的结果及评估	51
4.5	小结	52
第五章	分布式模糊概念格模型及构造算法	53
5.1	引言	53
5.2	分布式模糊概念格模型	53
5.3	基于模糊概念子格的并运算	54
5.4	基于分布式构造模糊概念格的模糊参数计算	56
5.5	分布式构造模糊概念格算法及实验	58
5.5.1	基本思想	58
5.5.2	并运算算法伪码	60

5.5.3	分布式构造模糊概念格算法伪码·····	60
5.5.4	实验及分析·····	62
5.6	小结·····	64
第六章	模糊概念格在聚类中的应用 ·····	66
6.1	基于模糊概念格的聚类·····	66
6.1.1	数据挖掘中的主要聚类方法·····	66
6.1.2	Pollant 的基于对象相似度的聚类方法·····	68
6.1.3	基于概念距离的聚类方法·····	71
6.1.4	模糊本体生成·····	74
6.2	小结·····	76
第七章	结束语 ·····	77
7.1	论文主要贡献·····	77
7.2	进一步工作·····	78
图表附录	·····	79
参考文献	·····	81
致 谢	·····	94

第一章 绪 论

1.1 课题研究的的目的和意义

在实际中大量信息是模糊的、不确定的,模糊信息的表示和处理有很好的应用背景。本文将模糊引入概念格,针对以往模糊信息研究大多是基于离散的模糊集合的局限,定义了一个广义的模糊概念格模型,并研究了模糊信息的表示和处理以及在数据挖掘中的应用,包括模糊与概念格的结合、模糊概念格构造、模糊关联规则提取、模糊概念格的分布式构造、模糊与本体的结合、基于模糊概念格的聚类技术、模糊本体的生成方法等。

本论文工作属于国家自然科学基金项目“基于 Rough 集合理论的概念格结构模型研究(69985004)”、“分布式概念格数学模型与算法研究(60275022)”和“面向本体的形式概念分析扩展模型与算(60575035)”中的内容,并得到上海市高等学校青年发展基金(03AQ99)的资助。

1.2 国内外研究概况

随着计算机发展,人们处理的信息量与日俱增,在实际中不确定的、模糊的信息越来越多,如年龄为 30 岁的人可以划归青年人的集合,也可以划归中年人的集合。关于模糊信息的表示与处理有很实用的研究背景。美国人扎德最早提出了模糊集理论并应用于模糊控制中,提出了隶属度的概念。关于模糊信息的表示有 Karl Erich Wolff 的模糊信息表示法^[37]、Burusco 的 L_模糊集法^[38]、Girard 的模

糊量词集法^[39,40]等。在 Karl Erich Wolff 的模糊信息表示法中,属性用模糊语言变量值表示,语言变量值可构造成标度格,采用此格分类形式背景中的对象。Burusco 等从 L -模糊集合的形式背景构造格,采用 fix point 理论^[41],给出了一个计算格结构的方法。Girard 研究了由模糊量词描述的数据的概念格的构造问题,通过在数据类型定义中插入模糊量词集处理模糊信息。

在以往的数据挖掘研究中,基于的信息多是标准的、而非模糊的。概念格作为一种优良的形式化分析工具,在人工智能、知识发现等诸多领域都有广泛的应用。但目前,大量的研究也多是基于标准形式背景即二值背景的。本体的研究同样多是基于标准背景做的。而在实际中,存在大量的模糊信息。因而,对其表示和处理的研究具有很好的应用价值。关于模糊概念格的模型及其应用研究即是本文的主要内容。

1.3 论文的主要研究内容

论文的主要研究内容包括模糊信息的处理、模糊背景的约简、模糊概念格的构造、模糊规则提取、模糊概念格的分布式构造、生成模糊本体的方法。

在论文的第一章绪论中,我们指出了本文研究的目的和意义并介绍了国内外的研究概况。

在论文的第二章中,我们阐述了概念格和模糊集理论,包括概念格的基本定义、基于标准形式背景的概念格构造的描述,概念格构造的一般方法等。

在论文的第三章中,我们研究了将模糊集理论引入概念格,描述了模糊概念格模型及模糊概念格的构造算法。定义了两个模糊参数 E 和 δ ,分别表示概念节点中对象的平均隶属度和对象的隶属度值相对平均隶属度的偏离程度。

在论文的第四章中,我们研究了基于模糊概念格的模糊关联规

则的生成,在模糊概念格的节点级上,使用模糊参数 δ 值的限定,可以避免生成非健壮节点对、从而避免提取效果欠佳的规则;实验分析了基于模糊概念格提取关联规则的算法的时空复杂度。

在论文的第五章中,针对格的构造的巨大的时空复杂度问题,我们将分布处理的思想引入了模糊概念格,提出了分布式模糊概念格模型及模糊概念子格的并运算,提出并实验分析了模糊概念格的分布式构造算法。实验结果表明此算法比未采用分布处理的构造算法在时空复杂度上有很大改进。

在论文的第六章中,我们研究了模糊概念格在聚类上的应用,提出了采用模糊聚类技术和模糊参数从模糊概念格生成模糊聚类、构造模糊概念层次、进而采用映射构造模糊本体的方法,并给出了一个实例。

在论文的第七章结束语部分中,我们总结了本文的工作,提出了下一步的研究工作,例如概念格与粗糙集的结合等。